DMD Fast Forward. Published on April 25, 2006 as doi:10.1124/dmd.106.009613
This article has not been copyedited and formatted. The final version may differ from this version.

Glutathione S-Transferase Omega 1 and Omega 2 Pharmacogenomics

Baidehi Mukherjee, Oreste E. Salavaggione, Linda L. Pelleymounter, Irene Moon, Bruce W. Eckloff, Daniel J. Schaid, Eric D. Wieben and Richard M. Weinshilboum

Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics (BM, OES, LLP, IM, RMW), Department of Biochemistry and Molecular Biology (BWE, EDW), and Department of Health Sciences Research (DJS), Mayo Clinic College of Medicine, Rochester, MN 55905, U.S.A.

Running title: **GSTO Pharmacogenomics**

Abbreviations: GSTO, glutathione S-transferase omega; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; WT, wild type; UTR, untranslated region; FR, flanking region; ORF, open reading frame; AA, African-American; CA, Caucasian-American; HCA, Han Chinese American; MA, Mexican-American; RRL, rabbit reticulocyte lysate

Address correspondence to:
Richard M. Weinshilboum, M.D.
Division of Clinical Pharmacology
Department of Molecular Pharmacology and Experimental Therapeutics
Mayo Clinic College of Medicine
200 First Street SW
Telephone: (507) 284-2246
Fax: (507) 284-9111
E-mail: Weinshilboum.Richard@mayo.edu

Text pages: 24 Number of tables: 5 Number of figures: 6

Number of references: 52

Number of words (includes Abstract, Introduction and Discussion): 1364

ABSTRACT

Glutathione S-transferase omega 1 and omega 2 (GSTO1 and GSTO2) catalyze monomethyl arsenate reduction, the rate-limiting reaction in arsenic biotransformation. As a step toward pharmacogenomic studies of these phase II enzymes, we resequenced human *GSTO1* and *GSTO2* using DNA samples from four ethnic groups. We identified 31 and 66 polymorphisms in *GSTO1* and *GSTO2*, respectively, with 4 nonsynonymous coding single nucleotide polymorphisms (cSNPs) in each gene. There were striking variations among ethnic groups in polymorphism frequencies and types. Expression constructs were created for all eight nonsynonymous cSNPs as well as a deletion of codon 155 in *GSTO1*, and those constructs were used to transfect COS-1 cells. Quantitative Western blot analysis, after correction for transfection efficiency, showed a reduction in protein level of greater than 50% for the GSTO1 Tyr32 variant allozyme when compared with wild type (WT), while levels for the Asp140, Lys208, Val236 and codon 155 deletion variant constructs were similar to that of the WT. For GSTO2, the Tyr130 and Ile158 variant allozymes showed 50% and 84% reductions in levels of expression, respectively, when compared to WT, while the Ile41 and Asp142 allozymes displayed levels similar to that of WT GSTO2. Rabbit reticulocyte lysate (RRL) degradation studies showed that the GSTO1 Tyr32 and the GSTO2 Tyr130, Ile158 and Asp142/Ile158 variant allozymes were degraded more rapidly than were their respective WT allozymes. These observations raise the possibility of functionally significant pharmacogenomic variation in the expression and function of GSTO1 and GSTO2.

Introduction

The GSTs are members of a gene superfamily of multifunctional enzymes that catalyze the conjugation of glutathione with electrophilic substrates (Nebert and Vasiliou, 2004; Hayes et al., 2005; Yang et al., 2006). Glutathione conjugation plays an important role in the detoxification of many drugs, carcinogens and endogenous compounds (Mulder and Ouwerkerk-Mahadevan, 1997; Pool-Zobel et al., 2005; Yang et al., 2006). Genetic variation in GSTs has been reported to represent a risk factor for a variety of cancers (Landi, 2000; Strange et al., 2000; Strange et al., 2001). The cytosolic GSTs are made up of alpha (A), mu (M), omega (O), pi (P), sigma (S), theta (T) and zeta (Z) families (Strange et al., 2001). GSTO1 and GSTO2 are the most recently discovered cytosolic GSTs. Members of this family do not display activity with "classical" GST substrates such as 1-chloro-2,4-dinitrobenzene, dichloromethane, or ethacrynic acid, but they do exhibit several enzymatic characteristics of glutaredoxins (Board et al., 2000; Whitbread et al., 2003). Although most GSTs have a serine or tyrosine residue at the active site, the GSTOs have cysteine, as well as a 19 amino acid extension at the N-terminus that is not present in other cytosolic GSTs (Board et al., 2000). The genes encoding GSTO1 and GSTO2 map to chromosome 10 and are separated by approximately 1.5 kb (Whitbread et al., 2003).

Both GSTO1 and GSTO2 catalyze the reduction of monomethyl arsenate from an oxidation state of +V to + III (Zakharyan et al., 2001; Schmuck et al., 2005). The use of enzyme purified from rabbit liver played an important role in the identification of GST01 and GST02 as enzymes that can catalyze monomethyl arsenate reduction (Maiorino and Aposhian, 1985; Zakharyan et al., 1995; Zakharyan and Aposhian, 1999). This reduction reaction is the rate-limiting step in arsenic biotransformation in humans (Zakharyan et al., 2001). Although arsenic also undergoes methylation in humans, that conjugation reaction requires prior reduction. The

methylation of arsenic is catalyzed by a recently identified arsenic methyltransferase (AS3MT). We recently also resequenced and performed functional genomic studies with the human *AS3MT* gene (Wood et al., 2006).

Arsenic contamination of drinking water is a public health problem worldwide (Tchounwou et al., 1999; Ratnaike, 2003; Tchounwou et al., 2003; Minamoto et al., 2005). Chronic arsenic exposure can cause skin, bladder, kidney, liver and lung cancer (Chen et al., 1992; Hopenhayn-Rich et al., 1998; Paul et al., 2004), and levels of arsenic metabolites vary greatly in the urine of individuals drinking the same contaminated water (Vahter, 2000; Marnell et al., 2003). Genetic variation in GSTO1 and/or GSTO2 – enzymes that catalyze the rate-limiting step in arsenic biotransformation in humans – might explain a portion of that variation. Arsenic is also used as a therapeutic agent, and arsenic trioxide is used to treat a variety of neoplasms (Maeda et al., 2001; Maeda et al., 2004; Ivanov and Hei, 2005). Patients treated with this drug can display striking arsenic trioxide-induced toxicity (Soignet et al., 1998; Mathews et al., 2006), another situation in which knowledge of GSTO pharmacogenomics might prove helpful.

There have been previous pharmacogenetic studies of the GSTOs (Marnell et al., 2003; Whitbread et al., 2003; Yu et al., 2003), but there remains need for a systematic, comprehensive study of genetic variation in *GSTO1* and *GSTO2*, as well as studies of the functional implications of that variation. The experiments described subsequently represent an attempt to perform comprehensive studies of genetic variation in *GSTO1* and *GSTO2*, as well as functional genomic characterization of polymorphisms that alter the amino acid sequence of proteins encoded by variant GSTO alleles. We observed large ethnic variation in the presence of common genetic polymorphisms in *GSTO1* and *GSTO2*, and several of those polymorphisms had functional

implications. Finally, our studies of four different ethnic groups utilized DNA samples that were deposited in the Coriell Cell Repository by the National Institutes of Health (NIH) – samples that are available to the biomedical investigative community for use in future experiments.

MATERIALS AND METHODS

DNA Samples. DNA samples from 60 African-American (AA), 60 Caucasian-American (CA), 60 Han Chinese-American (HCA) and 60 Mexican-American (MA) subjects were obtained from the Coriell Cell Repository (Camden, NJ). These samples were one of the sample sets approved by the NIH Pharmacogenetics Research Network for use in this type of gene resequencing study (www.PharmGKB.org). They had been deposited by the National Institute of General Medical Sciences, and all of these subjects had provided written consent for the use of their anonymized DNA for experimental purpose. The present studies were reviewed and approved by the Mayo Clinic Institutional Review Board.

***GSTO1* and *GSTO2* Resequencing**. Twenty-two separate PCR amplifications were performed with each DNA sample, 8 for *GSTO1* and 14 for *GSTO2*. Exons, splice junctions and a portion of the 5′-flanking region (5′-FR) were amplified for both genes. Primers were designed to hybridize within introns, within 5′-FRs or within 3′-untranslated regions (3′-UTRs). This approach ensured that a GSTO processed pseudogene which maps to chromosome 3 (Whitbread et al., 2003) would not be amplified. The 5′-end of each primer used to amplify the *GSTO1* and *GSTO2* coding regions included an M13 tag to make it possible to use dye primer sequencing chemistry (Chadwick et al., 1996). Resequencing of the *GSTO2* 5′-FR and the intergenic area utilized dye terminator chemistry. The sequences of the primers used to perform the resequencing experiments are listed in **Supplementary Table 1**.

PCR amplifications were performed with AmpliTaq Gold DNA polymerase (Perkin-Elmer, Foster City, CA). The 50 μl reaction mixtures consisted of 1 U of DNA polymerase, 5 μl of a 10-fold diluted DNA sample (160-190 ng DNA), 10 pmol of each primer, 0.05 mM dNTP (Boehringer Mannheim, Indianapolis, IN) and 5 μl of 10X PCR buffer that contained 15 mM

$MgCl_2$ (Perkin Elmer). Amplifications were performed with a Perkin Elmer model 9700 thermal cycler. PCR cycling parameters involved a 12 min hot start at 94°C, 30 cycles at 94°C for 30 s, 30 s at the annealing temperature and a final 10 min extension at 72°C. Amplicons were sequenced on both strands in the Mayo Molecular Biology Core Facility with an ABI 377 DNA sequencer using BigDye™ (Perkin Elmer) dye primer-sequencing chemistry or, for the *GSTO2* 5′-FR and the intergenic area, dye terminator sequencing chemistry. The sequences of samples in which a SNP was observed only once or those with ambiguous chromatograms were verified by performing independent amplifications, followed by resequencing. Chromatograms were analyzed using the PolyPhred 3.0 (Nickerson et al., 1997) and Consed 8.0 (Gordon et al., 1998) programs from the University of Washington as well as Mutation Surveyor (Soft Genetics, LLC, State College, PA) (Bardelli et al., 2003). The GenBank accession number for the *GSTO1* and *GSTO2* reference sequences were NT_030059.12. Reference sequences for GSTO1 and GSTO2 cDNAs were NM_004832.1 and NM_183239.1, respectively.

GSTO1 and GSTO2 Transient Expression. WT GSTO1 and GSTO2 cDNA open reading frame (ORF) sequences were amplified from a pooled human liver cDNA library (Clontech, Mountain View, CA). Expression constructs for the nonsynonymous cSNPs observed during the resequencing studies were generated by site-directed mutagenesis using "circular PCR". Sequences of the site-directed mutagenesis primers are also listed in **Supplementary Table 1**. The site-directed mutagenesis amplification reactions contained 1 U of *Pfu* Turbo DNA polymerase (Stratagene, La Jolla, CA) in a final volume of 25 μl, 10 pmol of primers, 0.24 mM dNTPs and 2.5 μl of 10X PCR buffer. The sequences of expression constructs were confirmed by sequencing in both directions after cloning into the eukaryotic expression vector pCR3.1 (Invitrogen, Carlsbad, CA). WT and variant GSTO1 and GSTO2 expression constructs

were then transfected into COS-1 cells, together with pSV-β-galactosidase DNA (Promega) to make it possible to use β-galactosidase activity to correct for transfection efficiency. Transfections were performed with Lipofectamine 2000 (Invitrogen), and the cells were cultured for 48 h in Dulbecco's modified Eagles medium (DMEM) (Bio Whittaker, Walkersville, MD) with 10% fetal bovine serum (Clontech). They were then homogenized with a Polytron homogenizer (Brinkmann Instruments, Westbury, NY); the homogenates were centrifuged at 100,000 x g for 1 h; and supernatant preparations were stored at -80°C.

GSTO1 and GSTO2 Western Blot Analysis. Rabbit polyclonal antibodies were generated to GSTO1 and GSTO2 (Cocalico Biologicals, Reamstown, PA) using polypeptides that corresponded to amino acids 10 to 30 for each of the proteins, with an additional cysteine residue at the amino terminus. These synthetic polypeptides differed at two amino acids for the two GSTO isoforms and were conjugated to keyhole limpet hemocyanin prior to the immunization of rabbits. For quantitative Western blot analysis, cytosol preparations of COS-1 cells transfected with the expression constructs were loaded on a 12% polyacrylamide gel on the basis of β-galactosidase activity to correct for possible variation in transfection efficiency. After SDS-PAGE, the proteins were transferred to a nitrocellulose membrane that was probed with rabbit antiserum to the appropriate GSTO isoform, diluted 1:2000 with "blocking buffer". Goat anti-rabbit horseradish peroxidase, diluted 1:20,000, was used as the secondary antibody. Bound antibody was detected with the ECL Western blotting system (Amersham Pharmacia Biotech, Piscataway, NJ). The level of immunoreactive protein was quantitated using IP Gel Lab (Biosystemetica, Plymouth, UK).

GSTO1 and GSTO2 *in vitro* Translation and Degradation. Selected GSTO1 and GSTO2 expression constructs were also transcribed and translated *in vitro* in the presence of

^{35}S-methionine (1000 Ci/mmol, 2.5 mCi total activity) (Amersham Pharmacia Biotech) using the TNT® coupled rabbit reticulocyte lysate (RRL) system (Promega, Madison, WI). The reaction mixture used to generate radioactively labeled protein was incubated at 30°C for 90 min, and 5 µl aliquots were used to perform SDS-PAGE with 10% gels. Protein degradation studies were performed as described previously (Wang et al., 2003). Specifically, 50 µl of an adenosine 5′-triphosphate (ATP) generating system, 50 µl of "untreated" RRL and 10 µl of ^{35}S-methionine radioactively labeled GSTO allozyme were mixed. This mixture was incubated at 37°C, and 10 µl aliquots were removed at various time intervals, followed by SDS-PAGE. After electrophoresis, the gels were dried, exposed to X-ray film and the bands of radioactively labeled protein were quantified.

Data Analysis. Values for π, θ and Tajima's D were calculated as described by Tajima (Tajima, 1989), followed by correction for length. D′ values as a measure of linkage disequilibrium were calculated as described by Hartl and Clark (Hartl and Clark, 2000) and Hendrick (Hendrick, 2000) and were displayed graphically. Haplotype analysis was performed as described by Schaid et al (Schaid et al., 2002) using the E-M algorithm. Group mean values were compared using student's t-test. Degradation data were plotted using GraphPad Prism (GraphPad Software, San Diego, CA).

RESULTS

***GSTO1* and *GSTO2* Resequencing**. *GSTO1* and *GSTO2* were resequenced using DNA samples from 60 AA, 60 CA, 60 HCA and 60 MA subjects. For *GSTO1*, 1.5 x 10^6 bp of DNA was sequenced and analyzed on both strands for the 240 DNA samples studied. A total of 29 SNPs were observed (**Table 1** and **Fig. 1A**). With 60 DNA samples per ethnic group, giving 120 chromosomes per group, we had 90% power to detect a variant allele if the true population frequency of that variant was at least 2%. Therefore, we had adequate power to detect fairly common – but not rare – variants. Two 3 bp deletions were also present, one GGC deletion in intron 1 and an AGG deletion that spanned the 5′-splice donor site for exon 4. Specifically, the "AGG" was deleted from an AGGAGGT → AGGT sequence, resulting in the creation of a new splice donor site and the deletion of codon 155, as reported previously (Board et al., 2000). Twenty-four polymorphisms were present in AA, 12 in CA, 15 in HCA and 18 in MA subjects. Seven of these polymorphisms were specific to AA, 4 to CA and 3 to MA. Twenty-one of the polymorphisms in AA subjects had frequencies ≥ 1% and 10 had frequencies ≥ 10%. Comparable data for samples from CA subjects were 8 polymorphisms with frequencies ≥ 1% and 4 ≥ 10%, while for HCA subjects, 7 had frequencies of ≥ 1% and 4 had frequencies ≥ 10% and in MA subjects, 8 polymorphisms had frequencies ≥ 1% and 3 ≥ 10%. Four nonsynonymous cSNPs were observed in *GSTO1*. Polymorphisms that resulted in Cys32Tyr and Ala236Val alterations in the encoded amino acid sequence were observed only in CA and MA subjects, respectively, while SNPs that resulted in Ala140Asp and Glu208Lys changes in amino acid sequence were present in all four populations.

For *GSTO2*, approximately 3 x 10^6 bp of DNA was sequenced and analyzed – including the "intergenic" area located between *GSTO1* and *GSTO2*, i.e., sequence located between the

final *GSTO1* exon and the initial *GSTO2* exon (**Fig. 1**). A total of 66 SNPs were observed (**Table 1** and **Fig. 1B**). In AA, CA, HCA and MA subjects, 54, 19, 17 and 36 SNPs were present, respectively. Nineteen of these polymorphisms were specific to AA, 5 to CA, and 2 each were specific to HCA and MA subjects. Four nonsynonymous cSNPs were also observed in *GSTO2*. SNPs resulting in Val41Ile and Leu158Ile changes in encoded amino acids were seen only in AA and CA subjects, respectively. A polymorphism resulting in a Cys130Tyr change in amino acid sequence was present in both AA and HCA subjects, and a SNP resulting in an Asn142Asp change in amino acid sequence was observed in all four populations. All polymorphisms for both genes were in Hardy-Weinberg equilibrium ($P > 0.05$). Overall, we observed 15 novel polymorphisms for *GSTO1* and 43 for *GSTO2*, including 3 novel nonsynonymous SNPs for *GSTO2* (**Table 1**). Thirty-four of the SNPs that we observed were present in dbSNP and seven in the HapMap Phase I. All of the SNPs that we identified can be viewed on the PharmGKB website.

We also calculated nucleotide diversity, a measure of genetic variation, adjusted for the number of alleles studied, for both genes and for the intergenic region. Two standard measures of nucleotide diversity are π, average heterozygosity per site, and θ, a population mutation measure that is theoretically equal to the neutral mutation parameter (Fullerton et al., 2000). Values for Tajima's D, a test for the neutral mutation hypothesis (Fullerton et al., 2000), were also calculated (**Table 2**). Under conditions of neutrality, Tajima's D should equal 0. AA subjects showed greater nucleotide diversity for *GSTO1*, based on π and θ values, than did the other groups (**Table 2**). For *GSTO1*, the Tajima's D value for HCA subjects approached statistical significance. For *GSTO2*, none of the values for Tajima's D were statistically

significant. For the intergenic region, Tajima's D value for AA subjects was statistically significant, raising the possibility that this area may be undergoing selection.

***GSTO1* and *GSTO2* Linkage Disequilibrium and Haplotype Analysis**. Haplotypes are proving of increasing value for use in epidemiologic association studies (Drysdale et al., 2000; Brodde and Leineweber, 2005). Therefore, we performed both pairwise studies of linkage disequilibrium and haplotype analysis with our gene resequencing data. For the linkage disequilibrium analysis, pairwise calculations of D′ values were performed for all polymorphisms. |D′| values can range from 0 to 1.0, with a value of 1.0 indicating that two polymorphisms are maximally associated, while 0 indicates they are randomly associated (Hartl and Clark, 2000; Hendrick, 2000). Graphical representations of statistically significant pairwise population-specific |D′| values for all *GSTO1* and *GSTO2* polymorphisms are shown in **Fig. 2**. Since *GSTO1* and *GSTO2* are separated by only about 1.5 kb, it is not surprising that we observed linkage disequilibrium for polymorphisms located in the two genes. Patterns for possible haplotype blocks were not well defined in the CA and HCA DNA, but they appeared to be similar for the AA and MA samples (**Fig. 2**).

Haplotype is the combination of alleles on a single chromosome or, in a more restricted definition, all polymorphisms present on a single allele (Altshuler et al., 2005). We also analyzed haplotypes for *GSTO1* and *GSTO2*. It is possible to determine a haplotype unequivocally if not more than one polymorphism per allele is heterozygous, but haplotypes can also be inferred computationally (Schaid et al., 2002). *GSTO1* and *GSTO2* haplotypes, both unequivocal and inferred, with frequencies ≥ 1% in at least one ethnic group, are listed in **Tables 3** and **4**. Haplotype designations were made on the basis of the amino acid sequence of the encoded allozyme. For example, the “WT”, the most common amino acid sequence in AA

subjects, was designated **1*. Letter designations were then added based on descending frequencies of haplotype encoding that amino acid sequence – beginning with the AA population. Therefore,**1A* was more frequent than **1B*, and **1B* was more frequent than **1C*, etc. For *GSTO1*, **2, *3, *4* and **5* designations were assigned to the Tyr32, Asp140, Lys208 and Val236 variants, respectively. The **4* haplotype with the Lys208 alteration in encoded amino acid sequence also included the deletion of codon 155 as a result of the 3 bp deletion at the exon 4-intron 4 splice donor site since these two variants always occurred together. The *6 haplotype included two changes in encoded amino acid sequence, plus the loss of codon 155. For *GSTO2, *2, *3,* and **4* designations were assigned to encoded allozymes that included Ile41, Tyr130 and Asp142 changes in amino acids, respectively. The *GSTO2*5* designation was used for an Asp142, Ile158 "double variant" that existed on the basis of an inferred haplotype.

Western Blot Analysis of GSTO1 and GSTO2. As an initial step in the study of the possible functional implications of nonsynonymous cSNPs in *GSTO1* and *GSTO2*, expression constructs were created for all variant allozymes – including the *GSTO1* codon 155 deletion – and those constructs were used to transfect COS-1 cells. For *GSTO1*, the codon 155 deletion was linked to a polymorphism that resulted in a Glu208Lys alteration in encoded amino acid; so a construct with both the codon 155 deletion and the Lys208 variant amino acid was also created. A mammalian cell line was transfected to ensure the presence of mammalian post-translational modification and the mammalian machinery for protein degradation. The GSTO expression constructs were cotransfected with β-galactosidase to make it possible to correct for variations in transfection efficiency. The antibodies for both GSTO1 and GSTO2 were directed against peptides that did not include any of the genetically variant amino acids. When GSTO1 was expressed in COS-1 cells, the Tyr32 variant allozyme displayed a striking decrease in the

quantity of expressed immunoreactive protein and the Val236 allozyme showed a significant increase, while values for the other constructs – including those with the codon 155 deletion – were similar to those for the WT protein (**Fig. 3** and **Table 5**). It should be pointed out that the Cys32Tyr polymorphism resulted in the loss of the critical GSTO1 active site cysteine (Board et al., 2000). When GSTO2 constructs were transiently expressed in COS-1 cells, levels of Tyr130, Ile158 and the double variant Asp142/Ile158 allozyme proteins were strikingly reduced, while the Ile41 and Asp142 allozymes were expressed at approximately 80% of the level of the WT allozyme (**Fig. 4** and **Table 5**). The GSTO2 antibody cross-reacted with a protein in the COS-1 cells with a molecular mass slightly higher than that for recombinant GSTO2 (**Fig 4**). We attempted to complement these expression studies with parallel assays of level of enzyme activity, but the assays currently available for GSTO1 and GSTO2 activities – assays that have been used to study purified, bacterially expressed protein (Schmuck et al., 2005) – lacked adequate sensitivity for application to these cultured mammalian cell cytosol preparations.

GSTO1 and GSTO2 Degradation Studies. Studies of a series of genetic polymorphisms and mutations have shown that the change of only one or two amino acids can alter the quantity of encoded protein during expression in mammalian cells (Thomae et al., 2002; Adjei et al., 2003; Thomae et al., 2003; Wang et al., 2003; Ji et al., 2005; Salavaggione et al., 2005; Martin et al., 2006). A common mechanism responsible for this effect is accelerated protein degradation (Wang et al., 2003; Weinshilboum and Wang, 2004). In an attempt to study mechanisms responsible for the striking decrease in expression of the GSTO1 Tyr32 and the GSTO2 Tyr130 and Ile158 variant allozymes, protein degradation studies were performed using the RRL. The RRL is commonly used to perform this type of study (Wang et al., 2003). Specifically, radioactive WT and variant allozymes were synthesized using "treated" RRL (**Figs.**

5A and **6A**), and the recombinant allozymes were then incubated for various periods of time with "untreated" RRL and an ATP generating system to follow the time course of protein degradation. The Tyr32 GSTO1 variant allozyme was degraded much more rapidly than was the WT allozyme (**Fig. 5B** and **5C**). For example, only 20% of Tyr32 protein remained after 24 h, while 60% of the WT allozyme remained (**Fig. 5C**). For GSTO2, the Tyr130, Ile158 and Asp142/Ile158 allozymes were also degraded much more rapidly than was the WT, with Ile158 and Asp142/Ile158 being degraded even more rapidly than was Tyr130 (**Fig. 6B** and **6C**). After 24 h, 26% of Tyr130, 6.6% of Ile158, 5.6% of Asp142/Ile158 and 73% of the WT protein remained (**Fig. 6C**). Therefore, accelerated degradation appeared to be responsible, at least in part, for differences in levels of the GSTO1 and GSTO2 variant allozymes after expression in COS-1 cells.

DISCUSSION

These studies were designed to determine the nature and extent of common variation in the genes encoding human GSTO1 and GSTO2 and to characterize the functional implications of that variation. We observed a total of 31 polymorphisms in *GSTO1*, 15 of which were novel, while *GSTO2* resequencing revealed a total of 66 polymorphisms, 43 of which were novel (**Fig. 1** and **Table 1**). Both genes had four nonsynonymous cSNPs and *GSTO1* also included the deletion of codon 155 – as described previously (Whitbread et al., 2003). A large number of haplotypes were also observed for both genes (**Tables 3** and **4**). For the initial functional genomic studies, expression constructs for GSTO1 and GSTO2 variant allozymes were used to transfect COS-1 cells and quantitative Western blot analyses were performed. Variant allozymes for each gene showed striking decreases in levels of expressed protein (**Figs. 3** and **4**), due, at least in part, to accelerated degradation (**Figs. 5** and **6**). Studies of genetic polymorphisms and mutations in other phase II drug-metabolizing enzymes have provided similar observations, most often as a result of proteasome-mediated degradation (Thomae et al., 2003; Wang et al., 2003; Ji et al., 2005) or, on the basis of recent data, as a result of intracellular protein aggregation and aggresome formation (Wang et al., 2005). Obviously, the implications of these observations for *in vivo* function must be pursued in the course of future studies. Since the rate-limiting step in arsenic biotransformation in humans is reduction catalyzed, at least in part, by GSTO1 and GSTO2 (Zakharyan et al., 2001; Schmuck et al., 2005), these pharmacogenomic results – when combined with similar data for the human arsenic methyltransferase gene, *AS3MT* (Wood et al., 2006) – may help us to understand differences among individuals in the metabolism and effect of arsenic in humans.

In summary, we have identified a large number of novel *GSTO1* and *GSTO2* polymorphisms, as well as a series of haplotypes for both genes. Functional studies of common nonsynonymous cSNPs in these genes revealed striking differences in levels of protein expression and the mechanism responsible – at least in part – was rapid degradation of the variant allozyme. Striking variations among ethnic groups in allele and haplotype frequencies were also observed – as they have been for many other human genes (Adjei et al., 2003; Wang et al., 2003; Ji et al., 2005; Salavaggione et al., 2005; Martin et al., 2006; Wood et al., 2006). These results represent an important step toward future pharmacogenomic studies of these new members of the GST family of proteins. It will now be possible to utilize these results for *GSTO1* and *GSTO2*, together with similar data for the human arsenic methyltransferase (*AS3MT*) gene (Wood et al., 2006), to test the hypothesis that genetic variation in all or some of these genes might represent risk factors for arsenic-dependent carcinogenesis and/or variation in response to therapy with arsenic trioxide.

Acknowledgements

We thank Mrs. Luanne Wussow for her assistance with the preparation of this manuscript.

References

Adjei AA, Thomae BA, Prondzinski JL, Eckloff BW, Wieben ED and Weinshilboum RM (2003) Human estrogen sulfotransferase (SULT1E1) pharmacogenetics: gene resequencing and functional genomics. *Br J Pharmacol* **139:**1373-1382.

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P and Consortium. IH (2005) A haplotype map of the human genome. *Nature* **437:**1299-1320.

Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JK, Parmigiani G, Kinzler KW, Vogelstein B and Velculescu VE (2003) Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300:**949.

Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, Danley DE, Hoth LR, Griffor MC, Kamath AV, Rosner MH, Chrunyk BA, Perregaux DE, Gabel CA, Geoghegan KF and Pandit J (2000) Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem* **275:**24798-24806.

Brodde OE and Leineweber K (2005) Beta2-adrenoceptor gene polymorphisms. *Pharmacogenet Genomics* **15:**267-275.

Chadwick RB, Conrad MP, McGinnis MD, Johnston-Dow L, Spurgeon SL and Kronick MN (1996) Heterozygote and mutation detection by direct automated fluorescent DNA sequencing using a mutant Taq DNA polymerase. *BioTechniques* **20:**676-683.

Chen CJ, Chen CW, Wu MM and Kuo TL (1992) Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *Br J Cancer* **66:**888-892.

Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G and Liggett SB (2000) Complex promoter and coding region beta 2-adrenergic

receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* **97:**10483-10488.

Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, Salomaa V, E. V, Perola M, Boerwinkle E and Sing CF (2000) Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* **67:**881-900.

Gordon D, Abajian C and Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8:**195-202.

Hartl DL and Clark AG (2000) Organization of genetic variation, in: *Principles of Population Genetics. Chapter 3*, pp 95-107, Sinauer Associates, Inc., Sunderland, MA.

Hayes JD, Flanagan JU and Jowsey IR (2005) Glutathione transferases. *Annu Rev Pharmacol Toxicol* **45:**51-88.

Hendrick PW (2000) *Genetics of Populations.* Jones and Bartlett Publ. pp 396-405, Sudbury, MA.

Hopenhayn-Rich C, Biggs ML and Smith AH (1998) Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol* **27:**561-569.

Ivanov VN and Hei TK (2005) Combined treatment with EGFR inhibitors and arsenite upregulated apoptosis in human EGFR-positive melanomas: a role of suppression of the PI3K-AKT pathway. *Oncogene* **24:**616-626.

Ji Y, Salavaggione OE, Wang L, Adjei AA, Eckloff B, Wieben ED and Weinshilboum RM (2005) Human phenylethanolamine N-methyltransferase pharmacogenomics: gene resequencing and functional genomics. *J Neurochem* **95:**1766-1776.

Landi S (2000) Mammalian class theta GST and differential susceptibility to carcinogens: a review. *Mutat Res* **463:**247-283.

Maeda H, Hori S, Nishitoh H, Ichijo H, Ogawa O, Kakehi Y and Kakizuka A (2001) Tumor growth inhibition by arsenic trioxide (As2O3) in the orthotopic metastasis model of androgen-independent prostate cancer. *Cancer Res* **61:**5432-5440.

Maeda H, Hori S, Ohizumi H, Segawa T, Kakehi Y, Ogawa O and Kakizuka A (2004) Effective treatment of advanced solid tumors by the combination of arsenic trioxide and L-buthionine-sulfoximine. *Cell Death Differ* **11:**737-746.

Maiorino RM and Aposhian HV (1985) Dimercaptan metal-binding agents influence the biotransformation of arsenite in the rabbit. *Toxicol Appl Pharmacol* **77:**240-250.

Marnell LL, Garcia-Vargas GG, Chowdhury UK, Zakharyan RA, Walsh B, Avram MD, Kopplin MJ, Cebrian ME, Silbergeld EK and Aposhian HV (2003) Polymorphisms in the human monomethylarsonic acid (MMA V) reductase/hGSTO1 gene and changes in urinary arsenic profiles. *Chem Res Toxicol* **16:**1507-1513.

Martin YN, Salavaggione OE, Eckloff BW, Wieben ED, Schaid DJ and Weinshilboum RM (2006) Human methylenetetrahydrofolate reductase pharmacogenomics: gene resequencing and functional genomics. *Pharmacogenet Genomics* **16:**265-277.

Mathews V, George B, Lakshmi KM, A. V, Bajel A, Balasubramaniam P, Shaji RV, Srivastava VM, Srivastava A and Chandy M (2006) Single agent arsenic trioxide in the treatment of newly diagnosed acute promyelocytic leukemia: durable remissions with minimal toxicity. *Blood* **107:**2627-2632.

Minamoto K, Mascie-Taylor CG, Moji K, Karim E and Rahman M (2005) Arsenic-contaminated water and extent of acute childhood malnutrition (wasting) in rural bangladesh. *Environ Sci* **12:**283-292.

Mulder GJ and Ouwerkerk-Mahadevan S (1997) Modulation of glutathione conjugation in vivo: how to decrease glutathione conjugation in vivo or in intact cellular systems in vitro. *Chem Biol Interact* **105:**17-34.

Nebert DW and Vasiliou V (2004) Analysis of the glutathione S-transferase (GST) gene family. *Hum Genomics* **1:**460-464.

Nickerson DA, Tobe VO and Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25:**2745-2751.

Paul PC, Chattopadhyay A, Manna AK and Dutta SK (2004) Skin cancers in chronic arsenic toxicity--a study of predictive value of some proliferative markers. *Indian J Pathol Microbiol* **47:**206-209.

Pool-Zobel B, Veeriah S and Bohmer FD (2005) Modulation of xenobiotic metabolising enzymes by anticarcinogens-focus on glutathione S-transferases and their role as targets of dietary chemoprevention in colorectal carcinogenesis. *Mutat Res* **591:**74-92.

Ratnaike RN (2003) Acute and chronic arsenic toxicity. *Postgrad Med J* **79:**391-396.

Salavaggione OE, Wang L, Wiepert M, Yee VC and Weinshilboum RM (2005) Thiopurine S-methyltransferase pharmacogenetics: variant allele functional and comparative genomics. *Pharmacogenet Genomics* **15:**801-815.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM and Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* **70:**425-434.

Schmuck EM, Board PG, Whitbread AK, Tetlow N, Cavanaugh JA, Blackburn AC and Masoumi A (2005) Characterization of the monomethylarsonate reductase and dehydroascorbate reductase activities of Omega class glutathione transferase variants: implications for arsenic metabolism and the age-at-onset of Alzheimer's and Parkinson's diseases. *Pharmacogenet Genomics* **15:**493-501.

Soignet SL, Maslak P, Wang ZG, Jhanwar S, Calleja E, Dardashti LJ, Corso D, DeBlasio A, Gabrilove J, Scheinberg DA, Pandolfi PP and Warrell Jr RP (1998) Complete remission after treatment of acute promyelocytic leukemia with arsenic trioxide. *N Engl J Med* **339:**1341-1348.

Strange RC, Jones PW and Fryer AA (2000) Glutathione S-transferase: genetics and role in toxicology. *Toxicol Lett* **112-113:**357-363.

Strange RC, Spiteri MA, Ramachandran S and Fryer AA (2001) Glutathione-S-transferase family of enzymes. *Mutat Res* **482:**21-26.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:**585-595.

Tchounwou PB, Patlolla AK and Centeno JA (2003) Carcinogenic and systemic health effects associated with arsenic exposure--a critical review. *Toxicol Pathol* **31:**575-588.

Tchounwou PB, Wilson B and Ishaque A (1999) Important considerations in the development of public health advisories for arsenic and arsenic-containing compounds in drinking water. *Rev Environ Health* **14:**211-229.

Thomae BA, Eckloff BW, Freimuth RR, Wieben ED and Weinshilboum RM (2002) Human sulfotransferase SULT2A1 pharmacogenetics: genotype-to-phenotype studies. *The Pharmacogenomics J* **2:**48-56.

Thomae BA, Rifki OF, Theobald MA, Eckloff BW, Wieben ED and Weinshilboum RM (2003) Human catecholamine sulfotransferase (SULT1A3) pharmacogenetics: common functional genetic polymorphism. *J Neurochem* **87:**809-819.

Vahter M (2000) Genetic polymorphism in the biotransformation of inorganic arsenic and its role in toxicity. *Toxicol Lett* **112-113:**209-217.

Wang L, Nguyen TV, McLaughlin RW, Sikkink LA, Ramirez-Alvarado M and Weinshilboum RM (2005) Human thiopurine S-methyltransferase (TPMT) pharmacogenetics: variant misfolding and aggresome formation. *Proc Natl Acad Sci USA* **102:**9394-9399.

Wang L, Sullivan W, Toft D and Weinshilboum R (2003) Thiopurine S-methyltransferase pharmacogenetics: chaperone protein association and allozyme degradation. *Pharmacogenetics* **13:**555-564.

Weinshilboum R and Wang L (2004) Pharmacogenetics: inherited variation in amino acid sequence and altered protein quantity. *Clin Pharmacol Ther* **75:**253-258.

Whitbread AK, Tetlow N, Eyre HJ, Sutherland GR and Board PG (2003) Characterization of the human Omega class glutathione transferase genes and associated polymorphisms. *Pharmacogenetics* **13:**131-144.

Wood TC, Salavaggione OE, Mukherjee B, Wang L, Klumpp AF, Thomae BA, Eckloff BW, Schaid DJ, Wieben ED and Weinshilboum RM (2006) Human arsenic methyltransferase (AS3MT) pharmacogenetics: gene resequencing and functional genomics studies. *J Biol Chem* **281:**7364-7373.

Yang P, Ebbert JO, Sun Z and Weinshilboum RM (2006) A role of the glutathione metabolic pathway in lung cancer treatment and prognosis: a review. *J. Clin Oncol* **in press**.

Yu L, Kalla K, Guthrie E, Vidrine A and Klimecki WT (2003) Genetic variation in genes associated with arsenic metabolism: glutathione S-transferase omega 1-1 and purine nucleoside phosphorylase polymorphisms in European and indigenous Americans. *Environ Health Perspect* **111:**1421-1427.

Zakharyan R, Wu Y, Bogdan GM and Aposhian HV (1995) Enzymatic methylation of arsenic compounds: assay, partial purification, and properties of arsenite methyltransferase and monomethylarsonic acid methyltransferase of rabbit liver. *Chem Res Toxicol* **8:**1029-1038.

Zakharyan RA and Aposhian HV (1999) Enzymatic reduction of arsenic compounds in mammalian systems: the rate-limiting enzyme of rabbit liver arsenic biotransformation is MMA(V) reductase. *Chem Res Toxicol* **12:**1278-1283.

Zakharyan RA, Sampayo-Reyes A, Healy SM, Tsaprailis G, Board PG, Liebler DC and Aposhian HV (2001) Human monomethylarsonic acid (MMA(V)) reductase is a member of the glutathione-S-transferase superfamily. *Chem Res Toxicol* **14:**1051-1057.

Footnotes

The gene resequencing data described in this paper have been deposited in the NIH-funded database PharmGKB with submission numbers PS205018, PS205019, PS205020, PS205021, PS205022, PS205026, PS205027, PS205028 for *GSTO1* and PS204659, PS205029, PS205030, PS205031, PS205032, PS205033 for *GSTO2*.

Supported in part by National Institutes of Health (NIH) grants R01 GM28157 (BM, OES and RMW), R01 GM35720 (OES and RMW) and U01 GM61388, The Pharmacogenetics Research Network (OES, LLP, IM, BWE, DJS, EDW and RMW).

Figure Legends

Figure 1. Human *GSTO1* (A) and *GSTO2* (B) polymorphisms. Exons that encode the ORF (open reading frame) are represented as black rectangles while white rectangles represent untranslated regions (UTRs). Arrows represent polymorphisms, with colors indicating frequency. Amino acid alterations resulting from nonsynonymous SNPs are also indicated. I/D indicates insertion/deletion events.

Figure 2. Human *GSTO1* and *GSTO2* linkage disequilibrium in four ethnic groups. The values represented are pairwise |D′| values for each polymorphism pair. All values shown in color were statistically significant ($P < 0.05$). Numbers for individual polymorphisms are those listed in **Table 1**.

Figure 3. GSTO1 quantitative Western blot analysis. (A) Representative Western blots, in triplicate, for GSTO1 allozymes with single amino acid sequence alterations, loaded on the gel on the basis of β-galactosidase activity to correct for transfection efficiency. (B) Average levels of immunoreactive protein. Each bar represents the average of 6 independent transfections (mean ± SEM).

Figure 4. GSTO2 quantitative Western blot analysis. (A) Representative Western blots, in triplicate, for GSTO2 allozymes with single amino acid sequence alterations, loaded on the gel on the basis of β-galactosidase activity to correct for transfection efficiency. The antibody crossreacted with a protein in COS-1 cells with a molecular mass slightly greater than that of GSTO2. (B) Average levels of immunoreactive protein. Each bar represents the average of 6 independent transfections (mean ± SEM).

Figure 5. GSTO1 rabbit reticulocyte lysate (RRL) degradation studies. (A) Representative autoradiograph of ^{35}S-methionine radioactively labeled recombinant human WT and Tyr 32

GSTO1. (B) Representative autoradiograph for ^{35}S-methionine radioactively labeled recombinant GSTO1 WT and Tyr32 allozymes at different time points in the degradation experiments. (C) GSTO1 WT and Tyr32 allozyme protein remaining at each time point, expressed as a percentage of the basal level. Each point is the average ± SEM of 3 independent experiments. Tyr32 differed significantly from the WT at each point ($P < 0.005$).

Figure 6. GSTO2 rabbit reticulocyte lysate (RRL) degradation studies. (A) Representative autoradiograph of ^{35}S-methionine radioactively labeled recombinant WT, Tyr130, Ile158 and Asp142/158Ile GSTO2. (B) Representative autoradiograph for ^{35}S-methionine radioactively labeled recombinant human GSTO2 WT and Tyr130, Ile158 and Asp142/158Ile allozymes at different time points in the degradation experiments. (C) GSTO2 WT and variant allozymes remaining at each time point, expressed as a percentage of the basal level. Each point is the average ± SEM of 3 independent experiments. All points for variant allozymes differed significantly from the WT ($P < 0.005$).

Table 1

No.	Location	Nucleotide	Sequence Change	Amino Acid Change	Frequency of Variant Alleles				dbSNP
					African-American	*Caucasian-American*	*Han Chinese-American*	*Mexican-American*	
GSTO1 Polymorphism									
1	5'-FR	-1309	T→C		0.175	0.000	0.075	0.008	
2	5'-FR	-1276	G→A		0.000	0.000	0.000	0.008	
3	5'-FR	-1242	G→A		0.125	0.292	0.142	0.233	rs 2164624
4	5'-FR	-883	C→T		0.017	0.000	0.000	0.000	
5	5'-FR	-851	G→T		0.017	0.000	0.000	0.008	
6	5'-FR	-320	A→G		0.017	0.142	0.150	0.092	rs 1191975
7	5'-FR	-243	T→C		0.250	0.000	0.000	0.008	
8	5'-FR	-220	A→G		0.008	0.042	0.025	0.000	rs 628480
9	5'-FR	-206	A→T		0.042	0.000	0.000	0.000	
10	IVS 1	5	G→A		0.035	0.000	0.000	0.000	rs 629771
11	IVS 1	49-51	GGC Deletion		0.083	0.241	0.134	0.217	rs 11509435
12	Exon 2	42	G→T		0.000	0.008	0.000	0.000	
13	Exon 2	95	G→A	Cys(32)Tyr	0.000	0.009	0.000	0.000	
14	IVS 3	13	G→T		0.297	0.000	0.008	0.01	
15	IVS 3	-112	G→A		0.000	0.000	0.000	0.008	
16	IVS 3	-102	C→T		0.300	0.000	0.008	0.008	rs 7101149
17	Exon 4	419	C→A	Ala (140) Asp	0.117	0.263	0.142	0.233	rs 4925 *
18	Exon 4	464-466	AGG Deletion	Δ Codon 155	0.042	0.025	0.008	0.025	rs 11509437
19	IVS 4	28	A→G		0.008	0.033	0.025	0.000	rs 17884170
20	IVS 4	34	C→A		0.000	0.008	0.000	0.000	
21	IVS 4	-60	T→C		0.000	0.008	0.000	0.000	
22	IVS 5	42	G→T		0.283	0.000	0.008	0.008	rs 7906892
23	IVS 5	70	A→T		0.283	0.000	0.008	0.008	rs 7906559
24	IVS 5	88	T→C		0.017	0.000	0.000	0.000	
25	IVS 5	98	C→T		0.017	0.000	0.000	0.000	
26	IVS 5	186	A→G		0.017	0.000	0.000	0.000	
27	IVS 5	-100	A→G		0.300	0.000	0.008	0.008	
28	IVS 5	-55	T→A		0.008	0.000	0.000	0.000	
29	Exon 6	622	G→A	Glu (208) Lys	0.042	0.025	0.008	0.025	rs 11509438 *
30	Exon 6	707	C→T	Ala (236) Val	0.000	0.000	0.000	0.050	rs 11509439
31	3'-UTR	728	G→A		0.300	0.000	0.008	0.008	rs 7589

GSTO2 Polymorphisms

32	5'-FR	-2012-2013	Insertion of G	0.292	0.000	0.000	0.008	
33	5'-FR	-1785	T→G	0.050	0.000	0.000	0.000	rs 529232
34	5'-FR	-1771	G→C	0.008	0.000	0.000	0.000	
35	5'-FR	-1664	C→T	0.300	0.000	0.000	0.008	
36	5'-FR	-1598	C→G	0.300	0.000	0.000	0.008	
37	5'-FR	-1576	G→C	0.300	0.000	0.000	0.008	
38	5'-FR	-1566	G→A	0.300	0.000	0.000	0.008	
39	5'-FR	-1541	C→T	0.008	0.000	0.000	0.000	
40	5'-FR	-1357	C→A	0.300	0.000	0.000	0.008	rs 12264646
41	5'-FR	-1189	T→C	0.300	0.000	0.000	0.008	rs 10509769 *
42	5'-FR	-1179	G→T	0.008	0.000	0.000	0.000	
43	5'-FR	-1105	G→A	0.117	0.258	0.167	0.242	
44	5'-FR	-1102	T→G	0.217	0.683	0.767	0.692	rs 641071 *
45	5'-FR	-1009	C→T	0.300	0.000	0.000	0.008	rs 12264844 *
46	5'-FR	-781	A→C	0.283	0.000	0.000	0.008	
47	5'-FR	-730-731	Insertion of T	0.283	0.000	0.000	0.008	
48	5'-FR	-637	C→T	0.275	0.000	0.000	0.008	
49	5'-FR	-634	C→G	0.275	0.000	0.000	0.008	
50	Exon 1	-625	A→G	0.000	0.000	0.017	0.000	
51	Exon 1	-530	C→G	0.271	0.000	0.000	0.008	
52	Exon 1	-526	C→A	0.000	0.008	0.000	0.000	
53	Exon 1	-463	G→A	0.271	0.000	0.000	0.008	
54	Exon 1	-450	A→C	0.271	0.000	0.000	0.008	
55	Exon 1	-424	G→A	0.017	0.000	0.000	0.000	
56	Exon 1	-410	T→G	0.008	0.000	0.000	0.000	
57	Exon 1	-293	A→T	0.050	0.000	0.000	0.000	
58	Exon 1	-234	C→T	0.008	0.000	0.000	0.000	
59	IVS 1	52	A→G	0.008	0.000	0.000	0.000	
60	IVS 1	152	G→T	0.033	0.000	0.000	0.000	
61	IVS 1	310-311	Deletion of CG	0.008	0.000	0.000	0.000	
62	IVS 1	311	G→A	0.300	0.000	0.000	0.008	
63	IVS 1	333	T→C	0.008	0.000	0.000	0.000	
64	IVS 1	402	C→T	0.042	0.025	0.017	0.025	
65	IVS 1	426	T→C	0.050	0.000	0.000	0.000	
66	IVS 1	435	T→A	0.017	0.000	0.000	0.000	
67	IVS 1	477-478	Insertion of T	0.008	0.033	0.033	0.000	
68	IVS 1	1694	A→G	0.050	0.000	0.000	0.008	

69	IVS 1	2003	A→G		0.300	0.000	0.000	0.008	
70	IVS 1	2146	A→G		0.000	0.017	0.000	0.008	
71	IVS 1	2168	G→A		0.200	0.000	0.017	0.017	
72	IVS 1	-1131	C→T		0.417	0.025	0.017	0.025	
73	IVS 1	-185	A→G		0.060	0.000	0.000	0.000	
74	IVS 1	-19	T→A		0.000	0.000	0.008	0.000	
75	IVS 1	-12	G→A		0.008	0.000	0.000	0.008	
76	Exon 2	-183	A→G		0.108	0.242	0.175	0.183	
77	IVS 2	-16	C→T		0.008	0.000	0.000	0.000	
78	Exon 3	121	G→A	Val (41) Ile	0.009	0.000	0.000	0.000	
79	IVS 4	20	C→T		0.200	0.560	0.642	0.667	rs 157077 *
80	IVS 4	39	C→T		0.000	0.000	0.000	0.008	
81	IVS 4	-38	A→C		0.000	0.017	0.000	0.017	
82	Exon 5	389	G→A	Cys (130) Tyr	0.008	0.000	0.017	0.000	
83	Exon 5	424	G→A	Asn (142) Asp	0.233)	0.692	0.783	0.742	rs 156697 *
84	IVS 5	74	G→A		0.233	0.692	0.783	0.742	
85	IVS 5	170	T→C		0.000	0.008	0.000	0.000	
86	Exon 6	472	C→A	Leu (158) Ile	0.000	0.008	0.000	0.000	
87	IVS 6	-55	C→G		0.058	0.000	0.008	0.000	
88	IVS 6	-21	C→T		0.108	0.000	0.008	0.000	
89	Exon 7	591	G→T		0.167	0.033	0.133	0.017	rs 3758571
90	Exon 7	630	C→T		0.117	0.000	0.000	0.008	rs 157078
91	3'-FR	1038	C→T		0.175	0.033	0.133	0.017	rs 12219470
92	3'-FR	1080	C→T		0.000	0.008	0.000	0.008	
93	3'-FR	1243	A→G		0.000	0.017	0.000	0.000	
94	3'-FR	1343	T→G		0.125	0.000	0.000	0.000	rs 17116848
95	3'-FR	1615	C→T		0.000	0.008	0.000	0.000	
96	3'-FR	1649	T→C		0.000	0.000	0.000	0.008	
97	3'-FR	1724	C→T		0.008	0.000	0.000	0.000	

Table 1. Human *GSTO1* and *GSTO2* polymorphisms. Polymorphism locations, alterations in nucleotides and encoded amino acids as well as frequencies in the four ethnic groups studied are listed. IVS represents intervening sequence (intron), and FR is flanking region. The SNPs in exons and 5′-FR have been numbered on the basis of their locations in the cDNA, with the 'A' in the translation initiation codon assigned (+1). Negative numbers are located 5′, and positive numbers 3′ to this position. SNPs in introns have been numbered from exon-intron splice junctions, with (+1) assigned to the initial position (first nucleotide) in the 5′-splice donor site and (-1) assigned to the first nucleotide in the 3′-splice acceptor site. The final column indicates whether that polymorphism was present in dbSNP. An asterisk indicates SNPs reported in the HapMap, Phase I.

Table 2.

	$\pi \times 10^{-3}$	$\theta \times 10^{-3}$	Tajima's D	P value
GSTO1				
AA	1.85 ± 1.06	1.51 ± 0.54	0.64	0.54
CA	0.52 ± 0.39	0.92 ± 0.36	-1.06	0.30
HCA	0.30 ± 0.29	1.14 ± 0.43	-1.86	0.06
MA	0.54 ± 0.40	1.24 ± 0.46	-1.51	0.13
GSTO2				
AA	1.30 ± 0.70	1.40 ± 0.40	-0.37	0.72
CA	0.50 ± 0.30	0.40 ± 0.20	0.62	0.55
HCA	0.50 ± 0.30	0.60 ± 0.20	-0.67	0.52
MA	0.50 ± 0.30	0.80 ± 0.30	-1.30	0.20
Intergenic Region				
AA	3.80 ± 2.03	2.20 ± 0.73	1.94	0.05
CA	0.55 ± 0.44	0.25 ± 0.18	1.79	0.07
HCA	0.43 ± 0.37	0.25 ± 0.18	1.10	0.30
MA	0.66 ± 0.50	1.70 ± 0.60	-1.65	0.10

Table 2. Estimates of values for π, θ and Tajima's D for *GSTO1*, *GSTO2* and the intergenic region (i.e., the region between the final exon in *GSTO1* and the initial exon in *GSTO2*) in four ethnic groups. AA indicates African-Americans; CA is Caucasian-American; HCA is Han Chinese-American; and MA is Mexican-American. Values are parameter estimates mean ± SE. P values refer to Tajima's D.

Table 3.

	Allele	AA	CA	HCA	MA	5'-FR (-1309)	5'-FR (-1242)	5'-FR (-851)	5'-FR (-320)	5'-FR (-243)	5'-FR (-220)	5'-FR (-206)	IVS 1 (+5)	IVS 1 (+49)	Exon 2 (+42)	Exon 2 (+95)	IVS 3 (+13)	IVS 3 (-102)	Exon 4 (+419)	Exon 4 (+464)	IVS 4 (+28)	IVS 5 (+42)	IVS 5 (+70)	IVS 5 (-100)	Exon 6 (+622)	Exon 6 (+707)	3'-UTR (+728)
o	***1A***	0.23	0.52	0.64	0.6	T	G	G	A	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***1B***	0.18	-	-	0.01	**C**	G	G	A	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***1C***	0.14	-	-	-	T	G	G	A	**C**	A	A	G	C	G	G	**T**	**T**	C	G	A	**T**	**T**	**G**	G	C	**A**
o	***1D***	0.08	-	-	-	T	G	G	A	T	A	A	G	C	G	G	**T**	**T**	C	G	A	**T**	**T**	**G**	G	C	**A**
i	***1E***	0.04	-	-	-	T	G	G	A	T	A	A	**A**	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***1F***	0.03	-	-	-	T	G	G	A	**C**	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
i	***1G***	0.03	-	-	-	T	G	G	A	**C**	A	**T**	G	C	G	G	**T**	**T**	C	G	A	**T**	**T**	**G**	G	C	**A**
i	***1H***	0.01	-	-	-	T	G	**T**	A	**C**	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***1I***	-	0.14		0.06	T	G	G	**G**	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
i	***1J***	-	0.03	0.03	-	T	**A**	G	A	T	**G**	A	G	C	G	G	G	C	C	G	**G**	G	A	A	G	C	G
o	***1K***	-	-	0.02	-	T	**A**	G	A	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
i	***1L***	-	-	0.01	-	**C**	G	G	A	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***1M***	-	-	-	0.02	T	G	G	A	T	A	**T**	G	C	G	G	G	C	C	G	A	G	A	A	G	C	G
o	***2***		0.01			T	G	G	A	T	A	A	G	C	**T**	**A**	G	C	C	G	A	G	A	A	G	C	G
o	***3A***	0.09	0.24	0.09	0.17	T	**A**	G	A	T	A	A	G	**D**	G	G	G	C	**A**	G	A	G	A	A	G	C	G
i	***3B***	0.01	0.01	-	0.02	T	**A**	G	A	T	A	A	G	C	G	G	G	C	**A**	G	A	G	A	A	G	C	G
o	***3C***	-	-	0.05	-	**C**	G	G	A	T	A	A	G	**D**	G	G	G	C	**A**	G	A	G	A	A	G	C	G
i	***3D***	-	-	-	0.03	T	**A**	G	**G**	T	A	A	G	**D**	G	G	G	C	**A**	G	A	G	A	A	G	C	G
i	***4***	0.04	0.03	0.01	0.01	T	G	G	A	T	A	A	G	C	G	G	G	C	C	**D**	A	G	A	A	**A**	C	G
o	***5***	-	-	-	0.04	T	G	G	A	T	A	A	G	C	G	G	G	C	C	G	A	G	A	A	G	**T**	G
i	***6 (*4, *5)***	-	-	-	0.01	T	**A**	G	A	T	A	A	G	**D**	G	G	G	C	**A**	**D**	A	G	A	A	**A**	C	G

Table 3. Human *GSTO1* haplotypes with frequencies ≥ 1%. If a haplotype included a variant amino acid sequence, it was included in the table even if its frequency fell below the 1% cutoff. Nucleotide positions are numbered as described in the legend for **Table 1**. Variant nucleotides as compared to the "reference sequence", i.e., the most common sequence in AA subject, are highlighted as white type against a black background. The initial column lists "observed" (o) and inferred (i) haplotypes. "D" indicates a 3 bp deletion.

Table 4.

	Allele	AA	CA	HCA	MA	5'-FR (-2012)	5'-FR (-1785)	5'-FR (-1771)	5'-FR (-1664)	5'-FR (-1598)	5'-FR (-1576)	5'-FR (-1566)	5'-FR (-1357)	5'-FR (-1189)	5'-FR (-1105)	5'-FR (-1102)	5'-FR (-1009)	5'-FR (-781)	5'-FR (-730)	5'-FR (-637)	5'-FR (-634)	Exon 1 (-530)	Exon 1 (-463)	Exon 1 (-450)	Exon 1 (-293)	IVS 1 (+311)
i	***1A***	0.17	-	-	-	I	T	G	T	G	C	A	A	C	G	T	T	C	I	T	G	G	A	C	A	A
i	***1B***	0.14	-	-	0.01	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
o	***1C***	0.07	0.22	0.13	0.16	D	T	G	C	C	G	G	C	T	A	T	C	A	D	C	C	C	G	A	A	G
i	***1D***	0.05	-	-	-	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1E***	0.04	-	-	0.01	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1F***	0.03	-	-	-	D	C	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	T	G
i	***1G***	0.03	0.01	0.01	0.01	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1H***	0.03	0.01	0.01	0.01	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1I***	-	0.03	-	0.01	D	T	G	C	C	G	G	C	T	A	T	C	A	D	C	C	C	G	A	A	G
i	***1J***	0.03	-	-	-	I	T	G	T	G	C	A	A	C	G	T	T	C	I	T	G	G	A	C	A	A
i	***1K***	0.02	-	-	-	D	C	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	T	G
i	***1L***	0.02	-	-	-	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1M***	-	-	0.02	0.01	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***1N***	0.02	-	-	-	D	T	G	C	C	G	G	C	T	G	T	C	A	D	C	C	C	G	A	A	G
i	***2***	0.01	-	-	-	D	T	C	C	C	G	G	C	T	A	T	C	A	D	C	C	C	G	A	A	G
o	***3***	-	-	0.02	-	D	T	G	C	C	G	G	C	T	A	T	C	A	D	C	C	C	G	A	A	G
o	***4A***	0.14	0.5	0.55	0.61	D	T	G	C	C	G	G	C	T	G	G	C	A	D	C	C	C	G	A	A	G
o	***4B***	0.03	-	0.06	0	D	T	G	C	C	G	G	C	T	G	G	C	A	D	C	C	C	G	A	A	G
o	***4C***	0.03	0.1	0.12	0.07	D	T	G	C	C	G	G	C	T	G	G	C	A	D	C	C	C	G	A	A	G
i	***5***	-	0.01	-	-	D	T	G	C	C	G	G	C	T	G	G	C	A	D	C	C	C	G	A	A	G

	Allele	IVS 1 (+402)	IVS 1 (+426)	IVS 1 (+435)	IVS 1 (+1694)	IVS 1 (+2003)	IVS 1 (+2168)	IVS 1 (-1131)	IVS 1 (-185)	Exon 2 (-183)	Exon 3 (+121)	IVS 4 (+20)	Exon 5 (+389)	Exon 5 (+424)	IVS 5 (+74)	Exon 6 (+472)	IVS 6 (-21)	Exon 7 (+591)	Exon 7 (+630)	3'- FR (+1038)	3'-FR (+1343)
	1A	C	T	T	A	G	G	C	A	A	G	C	G	G	G	C	C	G	C	C	T
	1B	C	T	T	A	A	A	C	A	A	G	C	G	G	G	C	C	G	C	C	T
	1C	C	T	T	A	A	G	C	A	G	G	C	G	G	G	C	C	G	C	C	T
	1D	C	T	T	A	A	G	C	A	A	G	C	G	G	G	C	C	G	C	C	T
	1E	C	T	T	G	A	G	C	A	A	G	C	G	G	G	C	C	G	T	C	T
	1F	C	C	T	A	A	G	C	G	A	G	C	G	G	G	C	C	G	T	C	T
	1G	T	T	T	A	A	G	T	A	A	G	C	G	G	G	C	C	G	C	C	T
	1H	T	T	T	A	A	G	T	A	A	G	C	G	G	G	C	C	G	C	C	T
	1I	C	T	T	A	A	G	C	A	A	G	T	G	A	A	C	C	G	C	C	T
	1J	C	T	T	A	G	G	C	A	A	G	C	G	G	G	C	T	G	C	C	G
	1K	C	C	T	A	A	G	C	G	A	G	C	G	G	G	C	C	T	T	T	T
	1L	C	T	A	A	A	A	C	A	A	G	C	G	G	G	C	C	T	C	T	T
	1M	C	T	T	A	A	A	C	A	A	G	C	G	G	G	C	C	T	C	T	T
	1N	C	T	T	A	A	A	C	A	A	G	C	G	G	G	C	C	T	C	T	G
	2	C	T	T	A	A	G	C	A	G	A	C	G	G	G	C	C	T	C	T	T
	3	C	T	T	A	A	G	C	A	G	G	C	A	G	G	C	C	G	C	C	T
	4A	C	T	T	A	A	G	C	A	A	G	T	G	A	A	C	C	G	C	C	T
	4B	C	T	T	A	A	G	C	A	A	G	T	G	A	A	C	C	T	C	T	T
	4C	C	T	T	A	A	G	C	A	A	G	C	G	A	A	C	C	G	C	C	T
	5	C	T	T	A	A	G	C	A	A	G	C	G	A	A	A	C	G	C	C	T

Table 4. Human *GSTO2* haplotype with frequencies ≥ 1%. If a haplotype included a variant amino acid sequence, it was included in the table even if its frequency fell below the 1% cutoff. Nucleotide positions are numbered as described in the legend for **Table 1**. Variant nucleotides as compared to the "reference sequence", i.e., the most common sequence in AA subjects, are highlighted as white type against a black background. The initial column lists "observed" (o) and inferred (i) haplotypes.

Table 5.

GSTO1 and GSTO2 Allozyme Immunoreactive Protein Level	
Allozyme	**% WT**
GSTO1	
WT	100 ± 1.6
Tyr32	46.5 ± 3.3
Asp140	110 ± 4.5
Lys208	106 ± 4.9
Val236	136 ± 4.4
Δ Codon 155	126 ± 7.7
Δ Codon 155/Lys208	99.6 ± 7.3
Δ Codon 155/Asp140/Lys208	87.7 ± 2.3
GSTO2	
WT	100 ± 2.3
Ile41	82.8 ± 1.6
Tyr130	49.4 ± 2.8
Asp142	76.1 ± 3.0
Ile158	17.3 ± 1.7
Asp142/Ile158	15.2 ± 2.5

Table 5. GSTO1 and GSTO2 variant allozyme protein expression levels after the transient transfection of COS-1 cells. The values listed are mean ± SEM for 6 independent transfections.

Human *GSTO1* and *GSTO2* Polymorphisms

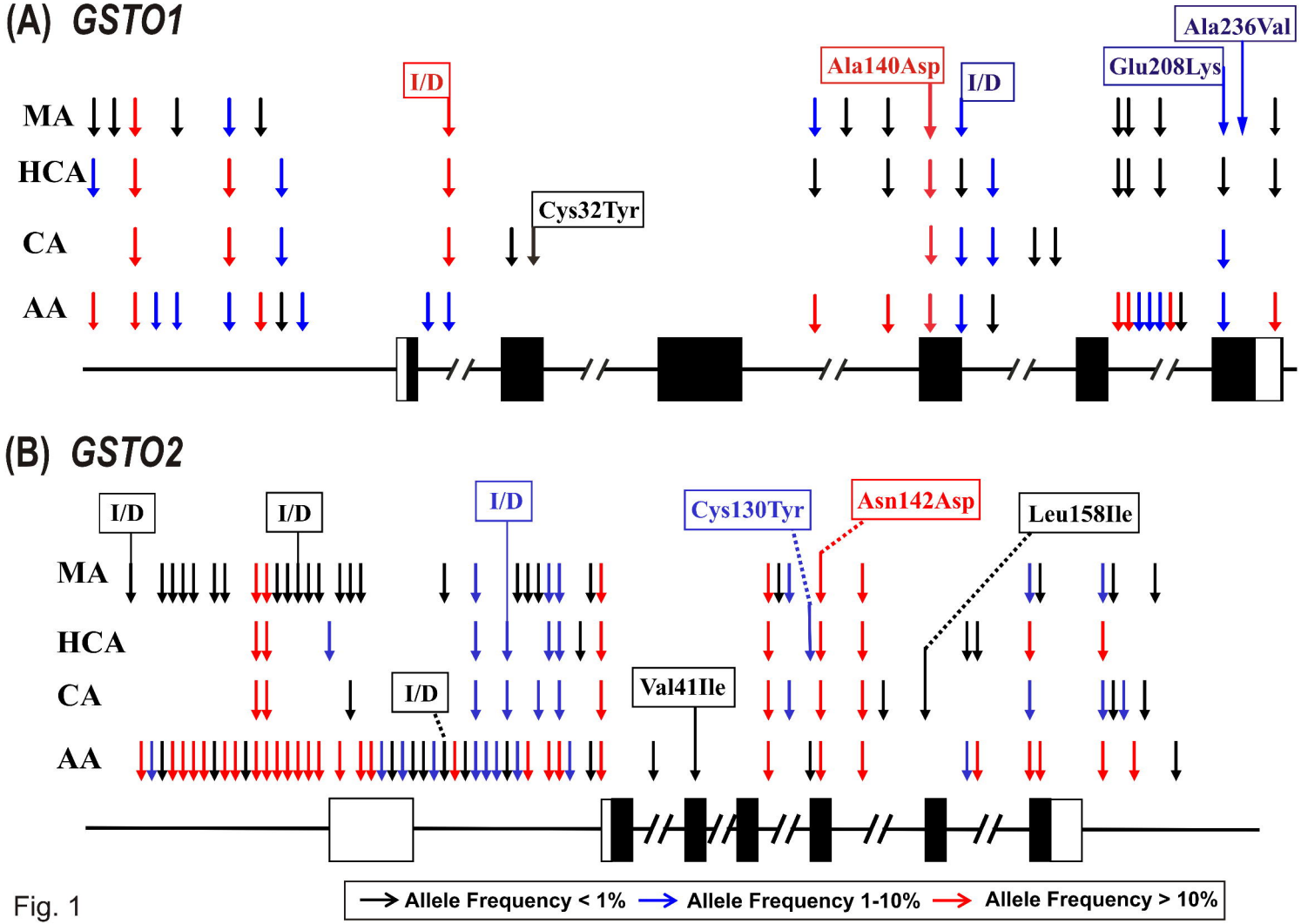

Fig. 1

Human *GSTO1* and *GSTO2* Linkage Disequilibrium

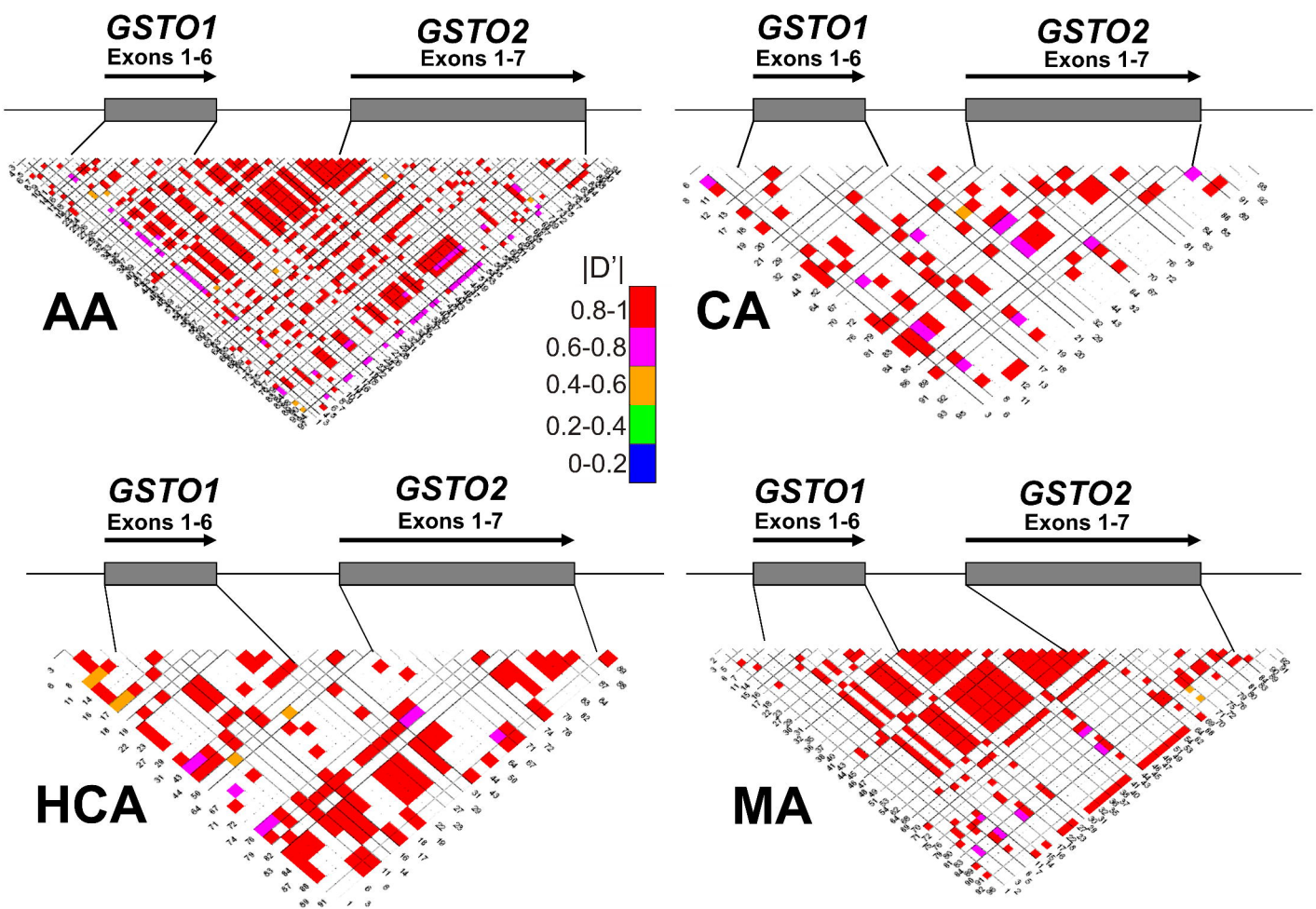

Fig. 2

Recombinant Human GSTO1 Allozymes

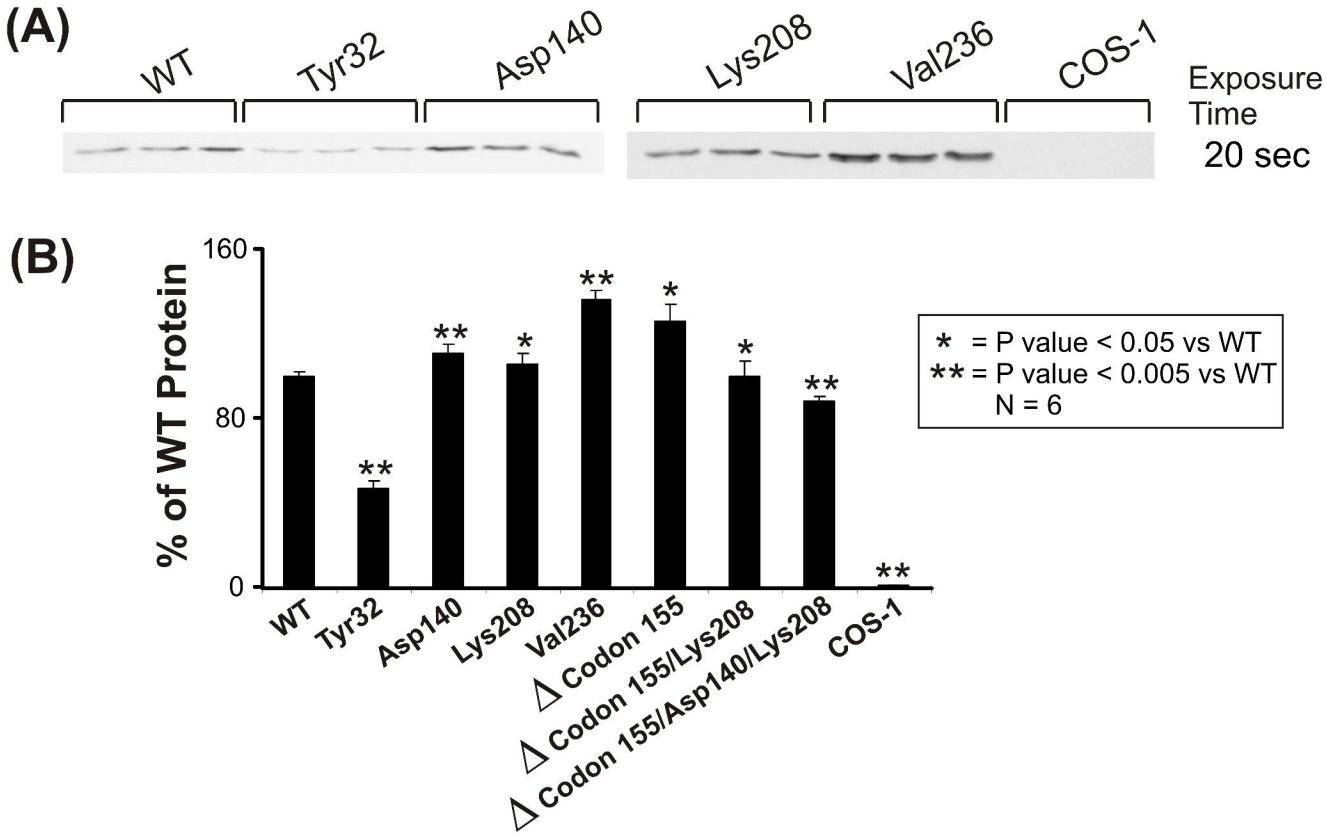

Fig. 3

Recombinant Human GSTO2 Allozymes

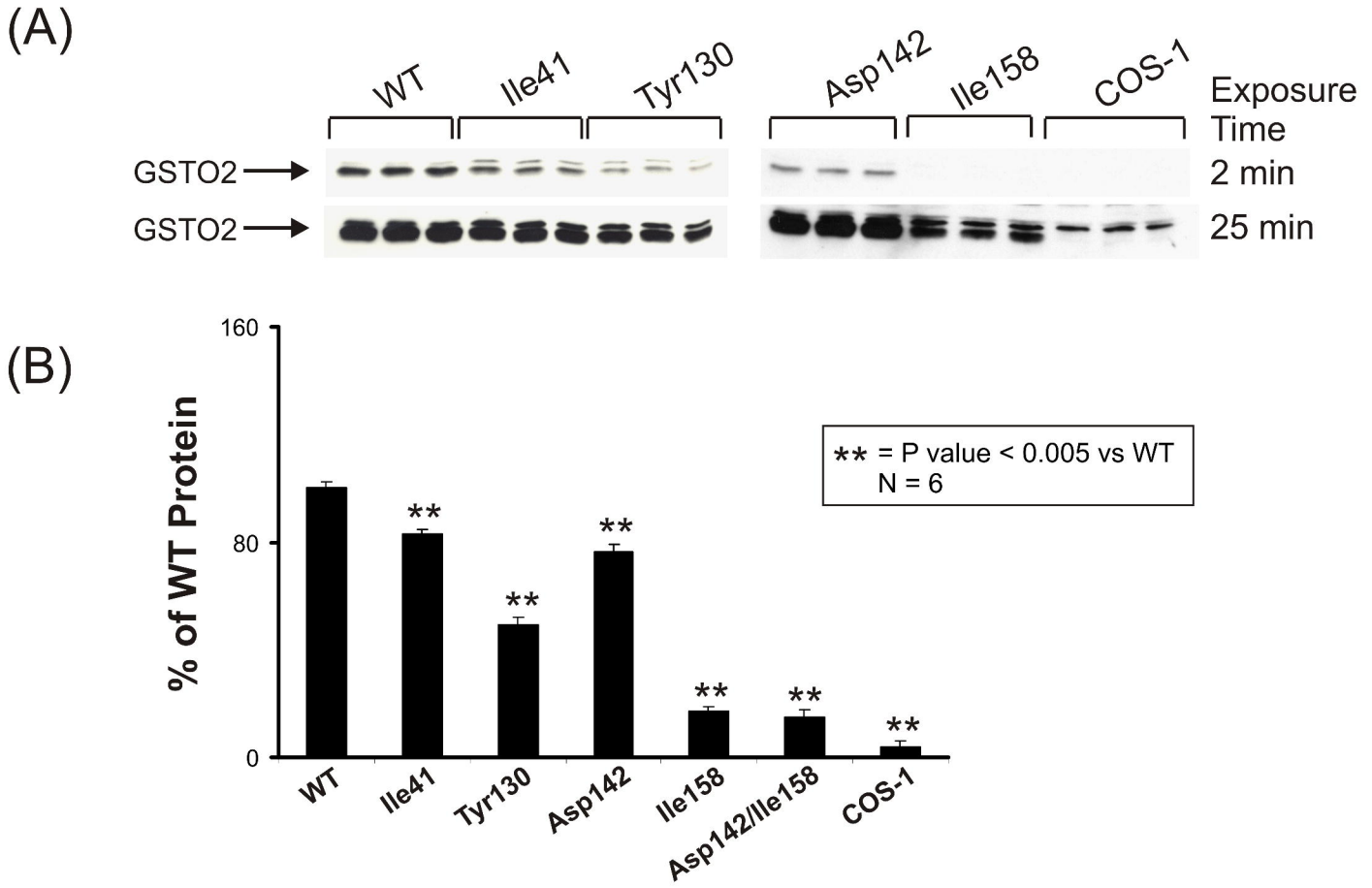

Fig. 4

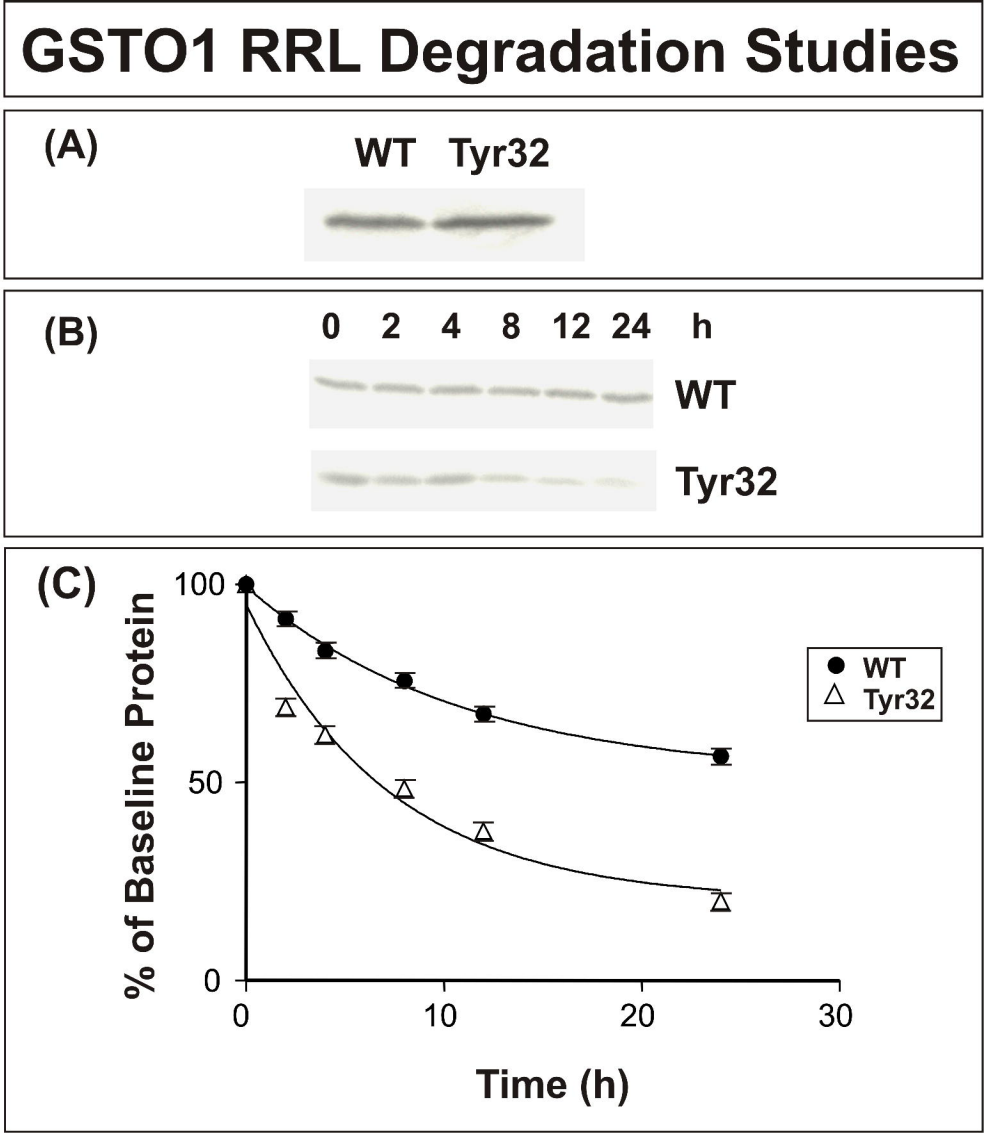
GSTO1 RRL Degradation Studies
(A)
WT
Tyr32
(B)
0
2
4
8
12
24
h
WT
Tyr32
(C)
% of Baseline Protein
100
50
0
0
10
20
30
Time (h)
WT
Tyr32

Fig. 5

GSTO2 RRL Degradation Studies

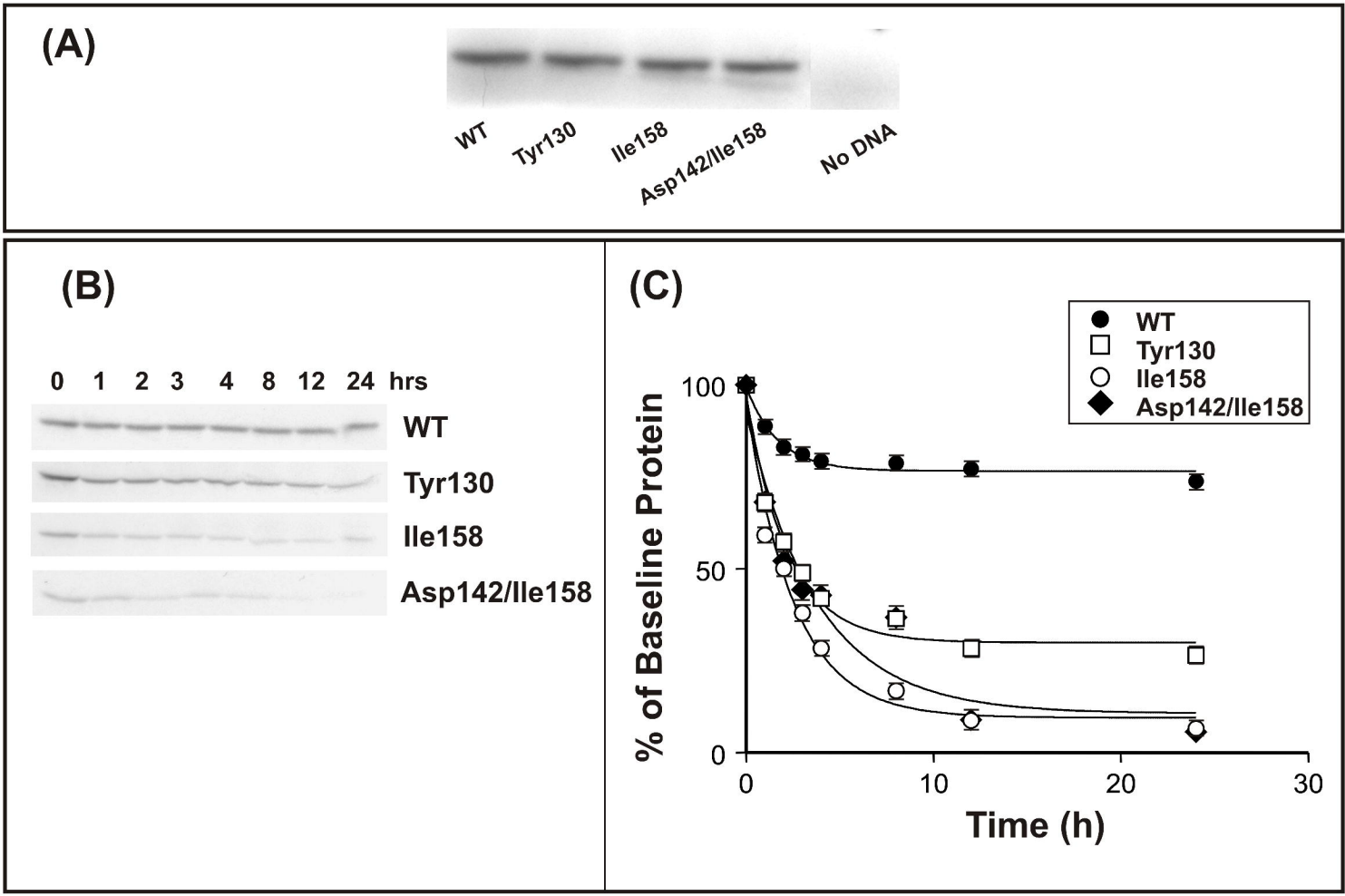

Fig. 6